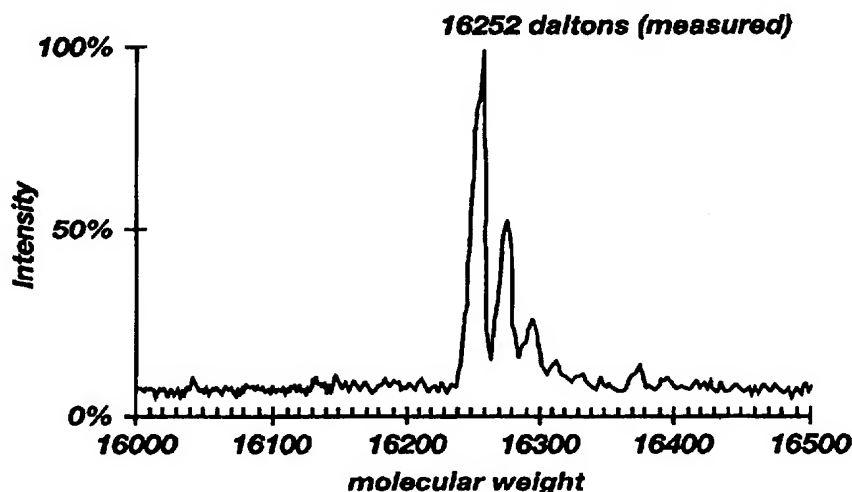




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12Q 1/68, C12P 19/34</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 97/47766</b> <b>(43) International Publication Date:</b> 18 December 1997 (18.12.97)
<b>(21) International Application Number:</b> PCT/US97/08518 <b>(22) International Filing Date:</b> 10 June 1997 (10.06.97)  <b>(30) Priority Data:</b> 60/019,702                      10 June 1996 (10.06.96)                      US  <b>(71) Applicant (for all designated States except US):</b> UNIVERSITY OF UTAH RESEARCH FOUNDATION [US/US]; 210 Park Building, Salt Lake City, UT 84112 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> HOPKINS, Chris, E. [US/US]; 103 Massasoit Road, Worcester, MA 01604 (US). GESTELAND, Raymond, F. [US/US]; 6877 Killyon Canyon, Salt Lake City, UT 84108 (US). PEIFFER, Andy [US/US]; 633 G. Street, Salt Lake City, UT 84103 (US). STAUFFER, Dora [US/US]; 1516 Hubbard Avenue, Salt Lake City, UT 84105 (US). LEPPERT, Mark [US/US]; 1466 Westminster Avenue, Salt Lake City, UT 84105 (US). CRAIN, Pamela, F. [US/US]; 2606 East 2100 South, Salt Lake City, UT 84109 (US). MCCLOSKEY, James, A. [US/US]; 2682 E. Eagle Way, Salt Lake City, UT 84108 (US).		<b>(74) Agents:</b> HOWARTH, Alan, J. et al.; Thorpe, North & Western, L.L.P., Suite 200, 9035 South 700 East, Sandy, UT 84070 (US).  <b>(81) Designated States:</b> CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

**(54) Title:** RAPID, ACCURATE IDENTIFICATION OF DNA SEQUENCE VARIANTS BY ELECTROSPRAY MASS SPECTROMETRY

**(57) Abstract**

A method of detecting a polymorphism at a polymorphic site in the genome of an individual is described. The method involves amplifying a portion of a genomic DNA sample by polymerase chain reaction to produce an amplified segment that contains the polymorphic site, desalting the amplified segment, and determining the mass of the amplified segment by electrospray ionization mass spectrometry. Comparison of the mass of the amplified segment to a reference mass permits detection of the presence or absence of the polymorphism. Methods for heterozygosity and disease inheritance by mass spectrometry are also described. The figure shows the molecular weight transform spectrum obtained from the mass-to-charge spectrum of a 53-base oligonucleotide.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RAPID, ACCURATE IDENTIFICATION OF DNA SEQUENCE  
VARIANTS BY ELECTROSPRAY MASS SPECTROMETRY

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S.  
5 Provisional Application No. 60/019,702, filed June 10,  
1996.

BACKGROUND OF THE INVENTION

This invention relates to detection of  
differences between nucleic acids. More particularly,  
10 the invention relates to rapid and accurate detection  
of differences between selected nucleic acids by  
electrospray mass spectrometry.

As the Human Genome Project matures, the demand  
for detecting genomic sequence differences is  
15 increasing. Many of these sequence differences, or  
polymorphisms, lead to altered gene expression, i.e.  
mutations. Analysis of the nucleotide composition for  
a polymorphism on both homologous chromosomes in the  
genome of an individual provides information relevant  
20 to detecting the inheritance of genetic diseases and  
disorders. Many disease mutations are thought to be  
caused by such small differences, either single base  
substitutions or single base insertions or deletions.  
Using polymerase chain reaction (PCR) technology, a  
25 segment of DNA containing a polymorphism can be  
amplified from each chromosome and then analyzed to  
determine nucleotide composition. Current methods of  
analysis rely heavily on electrophoretic methods of  
detection, where relative mobilities of DNA fragments  
30 are compared to extrapolate sequence differences.  
E.g., M. Orita et al., Detection of Polymorphisms of  
Human DNA by Gel Electrophoresis as Single-strand  
Conformation Polymorphisms, 86 Proc. Nat'l Acad. Sci.  
USA 2766-70 (1989); R.M. Meyers et al., Detection of  
35 Single Base Substitutions by Ribonuclease Cleavage at  
Mismatches in RNA:DNA Duplexes, 230 Science 1242-49

(1985); M.D. Traystman et al., Use of Denaturing Gradient Gel Electrophoresis to Detect Point Mutations in the Factor VIII Gene, 6 Genomics 293-301 (1990); S. Rust et al., Mutagenically Separated PCR (MS-PCR): A Highly Specific One-Step Procedure for Easy Mutation Detection, 21 Nucleic Acids Res. 3623-29 (1993); P.A.M. Roest et al., Protein Truncation Test (PTT) for Rapid Detection of Translation-termination Mutations, 2 Hum. Mol. Genet. 1719-21 (1993).

The most common technique for analysis of polymorphisms is nucleotide sequencing, particularly by the well-known dideoxynucleotide ("dideoxy") method. F. Sanger et al., DNA Sequencing with Chain-terminating Inhibitors, 74 Proc. Nat'l Acad. Sci. USA 5463 (1977). Sequencing a sample in which the locus for a polymorphism is identical for both chromosomes (a homozygote) generally provides a clean signal of the base composition for the polymorphic locus. On the other hand, if the nucleotide composition of a polymorphic locus is different on each chromosome (a heterozygote), the sequencing signal is complex. In base-substitution polymorphisms, sequencing generates a multiple base signal for the polymorphic position. In base-deletion or base-insertion polymorphisms, sequencing generates multiple base signals starting at the polymorphic locus and propagating throughout the rest of the sequence. Although accurate and reliable, dideoxy sequencing is complex and labor intensive requiring many hours to analyze a sample.

To simplify the electrophoretic analysis of polymorphisms, methods such as oligonucleotide ligation assay (OLA), mismatch PCR, and engineered restriction enzyme site analysis were developed. In OLA, two primers are designed to anneal adjacently on a genomic, amplified template. Subsequent thermocycling in the presence of a thermostable ligase covalently links the primers if they anneal to the

template without a gap between them and without a mismatch at the ends to be joined. This technique, and other similar techniques, is simple and about 5-times faster than sequencing, but it suffers from the drawback of requiring sequence-specific optimization to reduce false positives or false negatives. Another technique, sequence-based hybridization (SBH), where samples are analyzed based on their differential hybridization to stationary oligonucleotide arrays, requires less optimization of the sample preparation, but may require rearrangement of the oligonucleotide array to generate practical differential hybridization. The SBH technique is about 20-times faster than sequencing and is potentially simpler and thus more robust than the other techniques mentioned above. The speed of the electrophoretic and SBH techniques for analyzing polymorphisms appears to be approaching a minimum of minutes. To achieve polymorphism analysis methods that can be carried out in seconds, however, new methods of analysis are needed.

In view of the foregoing, it will be appreciated that providing a method for rapid and accurate detection and identification of sequence polymorphisms would be a significant advancement in the art.

#### BRIEF SUMMARY OF THE INVENTION

It is an object of the present invention to provide a rapid and accurate method of detecting polymorphisms in nucleic acids.

It is also an object of the invention to provide a method of detecting heterozygosity at a selected site in the genome of an individual.

It is another object of the invention to provide a method of detecting a genetic disorder or disease associated with a selected polymorphism in the genome of an individual.

These and other objects can be achieved providing a method of detecting inheritance by an individual of a selected genetic disease or disorder associated with a polymorphic site comprising a base substitution or a base deletion or insertion, comprising the steps of:

(a) collecting a genomic DNA sample from the individual;

(b) amplifying a selected portion of the genomic DNA sample by polymerase chain reaction to produce an amplified DNA segment containing the polymorphic site and having a size such that a difference in mass between the amplified DNA segment and a corresponding portion of a reference DNA sample can be resolved by electrospray ionization mass spectrometry;

(c) desalting the amplified DNA segment to produce a desalted amplified DNA segment;

(d) determining the mass of said desalted amplified DNA segment by electrospray ionization mass spectrometry; and

(e) comparing the mass to a reference mass determined for the corresponding portion of the reference DNA sample, wherein a difference between the mass and the reference mass indicates inheritance of the genetic disease or disorder.

While the size of the amplified DNA segment is to be limited only by functionality in detecting mass differences between the mass of the amplified DNA segment and the reference mass, the size of the amplified DNA segment is preferably no greater than about 55 base pairs, more preferably no greater than about 50 base pairs, and most preferably no greater than about 45 base pairs.

Various methods of desalting the amplified DNA can be used, as are known in the art. Preferred methods of desalting the amplified DNA include high pressure liquid chromatography, molecular weight cut-off spin filtration, and alcohol precipitation.

A method of detecting heterozygosity at a polymorphic site in the genome of an individual, wherein the polymorphic site comprises a site at which a base substitution or base deletion or insertion occurs in a mutant allele as compared to a wild type allele, comprises the steps of:

(a) collecting a genomic DNA sample from the individual;

(b) amplifying a selected portion of the genomic DNA sample by polymerase chain reaction to produce first and second amplified DNA segments representing corresponding portions of both homologous chromosomes containing the polymorphic site, wherein the first and second amplified DNA segments each have a size such that a difference in mass between the first and second amplified DNA segments can be resolved by electrospray ionization mass spectrometry;

(c) desalting the first and second amplified DNA segments to produce respective first and second desalted amplified DNA segments;

(d) determining the masses of the first and second desalted amplified DNA segments by electrospray ionization mass spectrometry; and

(e) comparing the masses of the first and second desalted amplified DNA segments, wherein a difference in the masses indicates heterozygosity at the polymorphic site.

A method of detecting a polymorphism at a polymorphic site comprising a base substitution or a base deletion or insertion, comprises the steps of:

(a) collecting a DNA sample to be tested;

(b) amplifying a selected portion of the DNA sample by polymerase chain reaction to produce an amplified DNA segment containing the polymorphic site and having a size such that a difference in mass between the amplified DNA segment and a corresponding

portion of a reference DNA sample can be resolved by electrospray ionization mass spectrometry;

(c) desalting the amplified DNA segment to produce a desalted amplified DNA segment;

(d) determining the mass of the desalted amplified DNA segment by electrospray ionization mass spectrometry; and

(e) comparing the mass to a reference mass determined for the corresponding portion of the reference DNA sample, wherein a difference between the mass and the reference mass indicates the presence of the polymorphism.

#### BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

FIG. 1 shows a mass-to-charge ( $m/z$ ) spectrum of a 53-base oligonucleotide (SEQ ID NO:1), wherein the values in brackets denote measured  $m/z$  values and the superscripts denote calculated charges.

FIG. 2 shows a molecular weight transform spectrum obtained from the mass-to-charge spectrum of FIG. 1.

FIG. 3 shows an  $m/z$  spectrum of PCR-amplified DNA.

FIG. 4 shows a molecular weight transform spectrum obtained from the  $m/z$  spectrum of FIG. 3.

FIG. 5 shows (A) a molecular weight transform spectrum obtained from PCR-amplified DNA wherein the template genomic DNA displayed homozygosity at the polymorphic site, and (B) a molecular weight transform spectrum obtained from PCR-amplified DNA wherein the template genomic DNA displayed heterozygosity at the polymorphic site.

FIG. 6 shows molecular weight transform spectra of pairwise mixtures of synthetic 53-residue oligonucleotides, wherein the paired oligonucleotides differ only by a single base, as follows: (A) C and A; (B) T and G; (C) A and G; (D) C and T; (E) C and G;



(F) T and A; (G) theoretical molecular weights of the four oligonucleotides.

FIG. 7 shows molecular weight transform spectra showing detection of alleles for a C to T base substitution polymorphism according to the present invention: (A) heterozygous mother; (B) heterozygous father; (C) homozygous daughter 1; (D) homozygous daughter 2; (E) theoretical molecular weights of amplified DNA strands from allele 1; (F) theoretical molecular weights of amplified DNA strands from allele 2.

FIG. 8 shows molecular weight transform spectra showing detection of alleles for a two-base deletion polymorphism according to the present invention: (A) heterozygous mother; (B) father homozygous for allele 1; (C) son 1 homozygous for allele 1; (D) heterozygous son 2; (E) theoretical molecular weights of amplified DNA strands from allele 1; (F) theoretical molecular weights of amplified DNA strands from allele 2.

#### DETAILED DESCRIPTION

Before the present method for rapid and accurate detection and identification of sequence polymorphisms is disclosed and described, it is to be understood that this invention is not limited to the particular configurations, process steps, and materials disclosed herein as such configurations, process steps, and materials may vary somewhat. It is also to be understood that the terminology employed herein is used for the purpose of describing particular embodiments only and is not intended to be limiting since the scope of the present invention will be limited only by the appended claims and equivalents thereof.

It must be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents

unless the context clearly dictates otherwise. Thus, for example, reference to "an amplified segment" includes reference to two or more of such amplified segments, reference to "a polymerase" includes  
5 reference to a mixture of two or more of such polymerases, and reference to "a primer" includes reference to two or more of such primers.

In describing and claiming the present invention, the following terminology will be used in accordance  
10 with the definitions set out below.

As used herein, "genomic DNA" or similar terms includes reference to transcription products of such genomic DNA. It is well known in the art that RNA can be readily copied as cDNA, which can be amplified by  
15 PCR. Thus, amplification of genomic DNA and amplification of cDNA obtained from RNA are considered equivalent.

Mass spectrometry provides a new method of DNA sequence analysis because differences in sequence  
20 composition can be measured from mass differences. Mass spectrometry is capable of very rapid sample analysis. For example, mass analysis of sequencing reaction products by electrospray ionization Fourier transform ion cyclotron resonance (ESI-FTICR) achieves  
25 mass analysis detection of all the reaction products with a 20-second sampling of the ion beam. Even though most mass spectrometric detectors are capable of determining the mass of components in a sample in less than a second, the rate for sample analysis is  
30 limited to about 15 minutes per sample due to other factors, such as signal averaging and sample handling. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF) has been  
35 demonstrated as a viable technique for analysis of polymorphisms. L.Y. Ch'ang et al., Detection of delta F508 mutation of the cystic fibrosis gene by matrix-assisted laser desorption/ionization mass

spectrometry, 9 Rapid Commun. Mass Spectrom. 772-74 (1995). The resolution limit for this MALDI-TOF analysis, however, is approached with detecting two-base deletion polymorphisms in double-stranded DNA. A recent advance with a modified form of MALDI-TOF, involving delayed extraction of the ion plasma, has increased its mass resolution somewhat. P. Juhasz et al., Applications of Delayed Extraction Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry to Oligonucleotide Analysis, 68 Anal. Chem. 941-46 (1996).

Mass spectrometric analysis has two stages, ionization and ion analysis. For analysis of polynucleotides, the ionization process employs either electrospray ionization (ESI) or Matrix-assisted laser desorption ionization (MALDI). The typical analysis hardware for either of these two ionizers can include ion trap (IT), time-of-flight (TOF), quadrupole, or Fourier transform ion cyclotron resonance (FTICR) analyzers. ESI is typically interfaced with a quadrupole analyzer, and MALDI is generally coupled with a TOF analyzer. The accuracy of mass analysis using these methods is approximately as shown in Table 1.

Table 1	
Ionizer-Analyzer	% Mass Error
ESI-Quadrupole	0.01
ESI-FTICR	0.0001
ESI-IT	0.1
MALDI-TOF	0.03-0.1
MALDI-FTICR	0.001

At the level of a 50-mer ( $M_r \sim 17,000$ ), all of these combinations of ionizer and analyzer should, in theory, be able to distinguish two molecules differing in mass by 15 daltons. However, in published results, TOF and IT analyzers both exhibit poor resolution such

that peaks differing by  $\pm 20$  daltons cannot be resolved. It has now been discovered that the peak resolution achieved for an ESI-quadrupole combination is capable of resolving 9 dalton peak differences, and that mass differences of 15 daltons are readily discernible.

There are six possible type of base substitution polymorphisms. Table 2 shows the absolute value of the mass differences for these six types of substitutions in double-stranded DNA.

Table 2				
	GC	AT	TA	CG
GC	0	1	1	0
AT		0	0	1
TA			0	1
CG				0

Table 3 shows the absolute value of the mass differences for these six types of substitutions in single-stranded DNA.

Table 3				
	G	A	T	C
G	0	16	25	40
A		0	9	24
T			0	15
C				0

Thus, for double-stranded DNA, the mass differences for these six types of base substitutions are only 0 or 1 dalton. For single-stranded DNA, however, the mass differences range from 9-40 daltons. Therefore, for base-substitution DNA polymorphisms to be analyzed by their mass differences, the samples must be denatured to single strands. Moreover, the separate mass analysis of each of the two strands of

double-stranded DNA provides two independent measurements of allelic composition. ESI-quadrupole analysis is ideally suited for analysis of base substitutions because it is inherently single-strand analysis. For a DNA sample to generate practical mass data by ESI-quadrupole analysis, it must be in a partially organic solvent that is virtually free of metallic, nonvolatile salts, e.g. methanol or acetonitrile. Thus, the low salt, organic nature of the solvent carrying a PCR product will generate, upon ionization, primarily single-stranded DNA ions.

An ESI-MS analyzer is capable of measuring very large ions because, as for all mass analyzers, the property measured is mass to charge ratio ( $m/z$ ), and not mass directly. Analysis of nucleic acids is generally performed in negative ion mode. As the nucleic acid enters the gas phase, the phosphate backbone oxyanions, all of which are charged in solution, are partially neutralized by proton adduction. The amount of neutralization of oxyanion charge varies between the molecules and thus results in a population of molecules with varying charge states by integral increments. Thus, mass analysis of a compound by mass spectrometry generates a series of  $m/z$  peaks, where the charge state of the peaks increases as the value of measured  $m/z$  decreases. For example, ESI-MS analysis of a synthetic oligonucleotide containing 53 nucleotide residues (SEQ ID NO:1) generates a series of  $m/z$  peaks (FIG. 1). For a given  $m/z$  peak shown in FIG. 1, the value inside the brackets is the measured  $m/z$ , and the superscript value is the calculated charge.

The mass of a compound can be calculated from one of the  $m/z$  peaks with the following formula:

$$M = (m/z)Z + Z,$$

where  $M$  is analyte mass,  $m/z$  is charge-to-mass ratio, and  $Z$  is the value of the charge. This formula is

used to calculate the charge from the measured value of two  $m/z$  peaks and their integral relationship of charge. Once charge is determined, the mass of the compound can be calculated from each  $m/z$  peak, and thus provide multiple measurements of molecular weight, as shown in Table 4. The average of the masses calculated from each  $m/z$  peak determines mass accuracy to within 0.01% mass error. Using the same formula, a computer can automatically transform an  $m/z$  spectrum into a molecular weight spectrum according to the Fenn method, J.B. Fenn et al., Electrospray Ionization for Mass Spectrometry of Large Biomolecules, 246 Science 64-70 (1989), hereby incorporated by reference, and thus render a molecular weight peak for each related  $m/z$  series, as shown in FIG. 2.

Table 4		
	$m/z$ Peak	Mass Estimate
772.9	21	16,252
811.6	20	16,252
854.3	19	16,251
901.9	18	16,252
954.9	17	16,250
1014.6	16	16,250
1082.5	15	16,253
1159.9	14	16,253
Average $\pm$ S.D.		16,251.6 $\pm$ 1.12

The presence of salt, e.g.  $\text{Na}^+$  or  $\text{K}^+$ , in the analyte solution competes with protons for neutralizing phosphate backbone charges. This salt adduction generates additional peaks in the ESI-MS spectrum. For example, in FIG. 2, the two peaks to the right of the main peak are salt adducts of  $\text{Na}^+$  and  $2 \text{Na}^+$ . If an ESI-MS analysis solution is inadequately desalted, then salt adduction peaks become the primary species observed. This is undesirable for at least

two reasons. First, the intensity of the signal is decreased by distributing the analyte mass over multiple peaks with varying amounts of salt adduction. Second, the mass spectrum of a complex mixture is confused by a much higher degree of  $m/z$  peak overlap.

To analyze polymorphisms in genomic material by ESI-quadrupole mass spectrometry, it is advantageous to use PCR technology. PCR primers are designed to flank a polymorphic locus such that the nucleotide base pairs involved in the polymorphism are not defined by the primers. Preferably, the distance between the primers is about 1-5 nucleotides. PCR amplification generates a dsDNA product from each chromosome of a homologous pair. In individuals homozygous for the polymorphic locus, where the specific nucleotide composition at the locus (allele) is the same for both homologous chromosomes, the PCR amplification will generate one type of dsDNA product. In heterozygous individuals, where the sequence composition at the locus is different for each chromosome (two different alleles), amplification generates two slightly different dsDNA products. Using ESI-quadrupole mass spectrometry, the molecule weight of each of the strands of a dsDNA PCR product can be measured. Each strand can be assigned to an allele type by comparison of the observed molecular weight to the expected weight of the allele types.

ESI-MS provides several advantages over methods known in the art. For example, the ESI-MS technique is highly specific. The small size of amplified products enables observation of very narrow windows into the genome. Only the region of DNA between the DNA primers, i.e. the window, represents the genomic DNA. In the PCR amplification, the primers define the polymerization of their opposite strand base pairs. Hence, if the annealed primers are separated by only one base pair at their 3' ends, then only that base

pair on each homologous chromosome, out of the 3 billion base pairs in the human genome, is viewed in the window. Larger windows could be designed, but may present problems. First, a larger window might incorporate multiple polymorphic loci, which could confuse or misidentify diagnosis. Second, a larger window increases the mass measured, but does not change the mass difference between alleles. Thus, larger mass amplification products decrease resolution of alleles in a polymorphism. Limiting the PCR amplification window to including only the bases involved in a putative polymorphism prevents interference from other nearby polymorphisms and maximizes resolution between alleles.

Unlike electrophoretic techniques, the ESI-MS technique does not require empirical optimization of analysis conditions to elucidate polymorphy. Once the PCR amplification has been optimized for specificity and yield, the preparation of samples for mass analysis is the same for any amplified product. As a result, successful ESI-MS analysis appears to be dependent primarily on sample cleanup and to be independent of nucleotide composition. Electrophoretic techniques such as SSCP, however, require extensive exploration of running conditions to detect specific polymorphisms. Mismatch PCR and polymorphism selective restriction site engineering both are site-specific, thus must be re-optimized for each locus analyzed.

The resolution limit of ESI-MS analysis for polymorphism detection is approached in T-to-A or A-to-T base-substitution polymorphisms. Resolution of the 9 dalton difference between such alleles, see Example 3 below, demonstrates that two peaks are resolvable as different alleles on a 53-nucleotide substrate using a quadrupole mass detector, but that there is considerable overlap of the peaks under such



conditions. Using a Fourier transform ion cyclotron resonance (FTICR) mass detector can result in about 100-fold increase in resolution of the alleles, thus making easy detection of alleles. Enhanced resolution of a T-to-A base-substitution polymorphism can be achieved on a quadrupole detector by substituting deoxyuridine for deoxythymidine in the PCR reaction protocol. This modification results in a 23-dalton separation between the alleles, thus enabling a quadrupole detector to readily distinguish the alleles of all types of base substitution polymorphisms in a 53-nucleotide substrate.

ESI-MS is complementary to current technology by enabling confirmation of putative polymorphic loci and then rapid screening of sample sets. The ESI-MS technique improves reliability of diagnosis because it directly measures the mass of the alleles involved in a polymorphism, not their relative mobilities. ESI-MS analysis appears well suited for detecting base-substitution polymorphisms or base deletion/insertion polymorphisms, which are difficult to detect by current methods. Many disease genes result from nonsense, missense, and frame-shifting mutations. These mutations, which are primarily caused by base substitutions, or one- or two-base deletions or insertions, are readily detectable by ESI-MS analysis.

ESI-MS can be used in the fields of disease gene detection, genotyping, tissue typing, and DNA forensics. ESI-MS analysis of a set of base substitution polymorphisms can be used to uniquely identify an individual, and thus may be a useful forensic tool. ESI-MS can also be used for tissue typing transplant or graft matching or for identifying pathogens, such as bacteria and viruses. This technique is also amenable to cancer diagnostics by detecting the presence of a somatically mutated allele in either tissue biopsies or in blood. ESI-MS

analysis can be used as a technique to confirm the content of polymorphisms by detecting the expected alleles and then to rapidly screen for the presence of the alleles in genomes.

5

## Example 1

10

15

20

25

30

35

Genomic templates were prepared by phenol/chloroform extraction of blood or of biopsy material according to methods well known in the art, e.g. J. Sambrook et al., Molecular Cloning: A Laboratory Manual (2d ed., 1989), hereby incorporated by reference. Oligonucleotides were either purchased from Genset or National Biosciences, or were synthesized according to methods well known in the art, e.g. S.A. Narang et al., 68 Meth. Enzymol. 90 (1979); E.L. Brown et al., 68 Meth. Enzymol. 109 (1979); U.S. Patent No. 4,356,270; U.S. Patent No. 4,458,066; U.S. Patent No. 4,416,988; U.S. Patent No. 4,293,652; N.D. Sinha et al., 24 Tetrahedron Lett. 5843 (1983); N.D. Sinha et al., 12 Nucl. Acids Res. 4539 (1984); N.D. Sinha et al., 15 Nucl. Acids Res. 397 (1987); N.D. Sinha et al., 16 Nucl. Acids Res. 319 (1988), hereby incorporated by reference. Taq polymerase was purchased from either Perkin-Elmer or as "PCR SUPERMIX" from Gibco/BRL. PCR was carried out in either a Model 9600 or a Model 2400 Perkin-Elmer thermocycler according to methods well known in the art, e.g. U.S. Patent No. 4,683,195; U.S. Patent No. 4,683,202; U.S. Patent No. 4,800,159; U.S. Patent No. 4,965,188; PCR Technology: Principles and Applications for DNA Amplification (H. Erlich ed., Stockton Press, New York, 1989); PCR Protocols: A guide to Methods and Applications (Innis et al. eds, Academic Press, San Diego, Calif., 1990), hereby incorporated by reference. PCR amplification was generally optimized at 40 cycles as follows: annealing for 30 seconds at 60-65°C, extension for 30 seconds at 72°C, and

denaturing for 10 minutes at 94°C. The primers were SEQ ID NO:2 and SEQ ID NO:3.

5 PCR product purification was performed by either reverse phase HPLC, molecular weight cut-off spin filtration, or ethanol precipitation. For most samples, the PCR products were isolated by reverse phase HPLC, lyophilized to dryness, resuspended in deionized water, and then ethanol precipitated. The supernate was then removed from the precipitate by aspiration, and the precipitate was lyophilized once again. The dry precipitate was then dissolved in 80% methanol, 10 mM triethylamine (TEA), and subjected to quadrupole mass analysis according to methods well known in the art.

15 Reverse phase HPLC was performed on a Waters HPLC with buffer A: 100 mM TEA-bicarbonate, pH 7, and buffer B: 100 mM TEA-bicarbonate, pH 7, and 50% methanol. Gradient elution was with 0 to 100% buffer B in 60 minutes, and the column was a 4 mm x 300 mm PRP-3 from Hamilton. Molecular weight cut-off spin filters were MICROCON 3 and MICROCON 10 from Amicon. Ethanol precipitation was performed at a final concentration of 70% ethanol and 0.7 M ammonium acetate at -20°C for greater than 1 hour.

25 Samples for ESI-MS analysis were solvated in 80% methanol, 10 mM TEA, and then were injected into a Sciex QE mass spectrometer through a charged capillary. Ions were measured as a mass-to-charge ratio spectrum and then transformed into molecular weight spectra by the Fenn method, *supra*.

30 FIGS. 3 and 4 show, respectively, the  $m/z$  spectrum and the molecular weight transform spectrum of PCR products amplified according to this example. Four major peaks were observed (FIG. 4), which are further characterized in Table 5.

Table 5			
Peak	Identity	Predicted Mass	Observed Mass
I	SEQ ID NO:1	16,211	16,211
II	SEQ ID NO:4	16,419	16,419
III	SEQ ID NO:5	16,524	16,523
IV	SEQ ID NO:6	16,731	16,732

These results show that two of the PCR products, peaks I and II, correspond to the sense and antisense strands of the predicted amplification product. The other two PCR products, peaks III and IV, correspond to the 3'-mono-adenylation products of peaks I and II. Taq polymerase is known to add such adenylate residues to amplification products in PCR. Table 5 also shows that comparison of the masses measured from the four amplification products, peaks I-IV, to the expected masses for such products demonstrates a mass-determining accuracy to within 0.01% mass error.

#### Example 2

In this example, the procedure of Example 1 was followed except that genomic templates were from an individual homozygous at a polymorphic site and from another individual heterozygous at the polymorphic site. SEQ ID NO:1 illustrates a portion of the sequence of the sense strand of one allele of the polymorphic site, and SEQ ID NO:7 illustrates a corresponding portion of the sequence of the sense strand sequence of the other allele at the polymorphic site. Thus, the two alleles differ only by a single base substitution. In the homozygote, where each homolog has the same allele, the mass analysis resulted in one peak for each of the four expected strands (FIG. 5A). In the heterozygote, however, where one homolog has a CG base pair at the allelic locus and the other homolog has a TA base pair, the

mass analysis generated doublet peaks for each strand (FIG. 5B). The molecular weight observed for each peak in the doublet corresponds to the expected molecular weight from each allele. In closer examination of the spectra, the spectrum from the homozygote shows a peak for the expected adenylated antisense strand, but the spectrum from the heterozygote does not appear to have an expected doublet at this position. The three other doublet peaks in the spectrum from the heterozygote each provide independent measurements of heterozygosity, thus the inherent redundancy of the technique enhances its robustness.

#### Example 3

In this example, oligonucleotides were synthesized to model all possible base substitution polymorphisms at the locus examined in Example 2. These four oligonucleotides were identical except for having a different base at the polymorphic site. The four oligonucleotides are SEQ ID NO:1, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9. Pairwise mixtures were made of all six possible pairs of oligonucleotides, and these mixtures were analyzed by ESI-MS according to the procedure of Example 1. FIG. 6 shows that 5 of the 6 pairs are easily resolved from each other (FIGS. 6A-E). The mixture of SEQ ID NO:7 and SEQ ID NO:8, having, respectively, T and A residues at the polymorphic site, shows partial overlap of the peaks (FIG. 6F). Better resolution could be obtained by decreasing the size of the oligonucleotides, which would increase the percentage difference in mass between the oligonucleotides.

#### Example 4

In this example, the procedure of Example 3 is followed except that SEQ ID NO:10 is substituted for

SEQ ID NO:7. The 23 dalton difference in molecular weight between dA and dU at the polymorphic locus results in the mixture of SEQ ID NO:8 and SEQ ID NO:10 being easily resolved.

#### Example 5

In this example, the utility of the present technique was explored for genotyping and for disease inheritance detection in family groups. ESI-MS genotyping in a four-member family for a C to T base substitution polymorphism was performed at a silent polymorphism locus in a candidate gene for Benign Neonatal Familial Convulsions (BNFC). The procedure was as in Example 1 except that the primers were SEQ ID NO:11 and SEQ ID NO:12. The molecular weight spectra (FIG. 7) demonstrate detection of alleles, namely two alleles in the heterozygotic mother (A) and father (B), and only one allele in the homozygotic daughters (C and D). The ESI-MS analysis and assignment were performed without knowledge of which samples were homozygotic or heterozygotic. The assignment was confirmed by checking the sequence results from an automated sequencer that had previously been used to diagnose allele inheritance in the family. In comparison to the sequencing results, ESI-MS analysis demonstrates at least two advantages. First, ESI-MS generates data on the composition of each strand, sense and antisense, in the analysis of one sample, whereas automated sequencing requires two separate sequencing reactions, i.e. with a forward primer and a reverse primer, to achieve data on each strand. Second, ESI-MS gives positive detection of each allele in heterozygotes, whereas automated sequencing gives complex signals that can be difficult or impossible to interpret without performing additional experimentation.

## Example 6

In this example, ESI-MS technique was applied to disease inheritance detection. Analysis was performed on a family group with a defective allele for the Attenuated Polyposis Coli (APC) gene. The defective allele comprises a two-base deletion, which results in a 618 dalton difference between the PCR products of the alleles. FIG. 8 shows spectra obtained according to the procedure of Example 1, except that the primers were SEQ ID NO:13 and SEQ ID NO:14. Spectra FIGS. 8A and 8D show the presence of peaks that are not present in the spectra FIGS. 8B and 8C, indicating the presence of the mutant allele.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) **APPLICANT:** Hopkins, Chris E.  
Gesteland, Raymond F.  
Peiffer, Andy  
Stauffer, Dora  
Leppert, Mark  
Crain, Pamela F.  
McCloskey, James A.

(ii) **TITLE OF INVENTION:** Rapid, Accurate  
Identification of DNA  
Sequence Variants by  
Electrospray Mass  
Spectrometry

(iii) **NUMBER OF SEQUENCES:** 14

(iv) **CORRESPONDENCE ADDRESS:**

(A) **ADDRESSEE:** Thorpe, North & Western,  
L.L.P.  
(B) **STREET:** 9035 South 700 East, Suite 200  
(C) **CITY:** Sandy  
(D) **STATE:** Utah  
(E) **COUNTRY:** USA  
(F) **ZIP:** 84070

(v) **COMPUTER READABLE FORM:**

(A) **MEDIUM TYPE:** Diskette, 3.5 inch, 1.44  
Mb storage  
(B) **COMPUTER:** Toshiba T2150CDS  
(C) **OPERATING SYSTEM:** Windows 95  
(D) **SOFTWARE:** Word Perfect 7.0

(vi) **CURRENT APPLICATION DATA:**

(A) **APPLICATION NUMBER:**  
(B) **FILING DATE:**  
(C) **CLASSIFICATION:**

(vii) **PRIOR APPLICATION DATA:**

(A) **APPLICATION NUMBER:** 60/019,702  
(B) **FILING DATE:** 10-JUN-1996

(viii) **ATTORNEY/AGENT INFORMATION:**

(A) **NAME:** Alan J. Howarth  
(B) **REGISTRATION NUMBER:** 36,553  
(C) **REFERENCE/DOCKET NUMBER:** T3262.PCT/U-



(ix) **TELECOMMUNICATION INFORMATION:**

- (A) **TELEPHONE:** (801)566-6633
- (B) **TELEFAX:** (801)566-0750

(2) **INFORMATION FOR SEQ ID NO:1:**5 (i) **SEQUENCE CHARACTERISTICS:**

- (A) **LENGTH:** 53 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

10 (xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:1:

CCTACAACAC CAGGAAGTAC GAGTGCTGCG CCGAGATCTA 40  
CCCGGACATC ACC 53

(2) **INFORMATION FOR SEQ ID NO:2:**15 (i) **SEQUENCE CHARACTERISTICS:**

- (A) **LENGTH:** 25 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:2:

20 CCTACAACAC CAGGAAGTAC GAGTG 25

(2) **INFORMATION FOR SEQ ID NO:3:**(i) **SEQUENCE CHARACTERISTICS:**

- (A) **LENGTH:** 24 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:3:

GGTGATGTCC GGGTAGATCT CGGC 24

(2) **INFORMATION FOR SEQ ID NO:4:**30 (i) **SEQUENCE CHARACTERISTICS:**

- (A) **LENGTH:** 53 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:4:

GGTGATGTCC GGGTAGATCT CGGCGCAGCA CTCGTACTTC 40

CTGGTGTTGT AGG 53

(2) **INFORMATION FOR SEQ ID NO:5:**

(i) **SEQUENCE CHARACTERISTICS:**

(A) **LENGTH:** 54 base pairs

(B) **TYPE:** nucleic acid

(C) **STRANDEDNESS:** single

(D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:5:

CCTACAACAC CAGGAAGTAC GAGTGCTGCG CCGAGATCTA 40

CCCGGACATC ACCA 54

(2) **INFORMATION FOR SEQ ID NO:6:**

(i) **SEQUENCE CHARACTERISTICS:**

(A) **LENGTH:** 54 base pairs

(B) **TYPE:** nucleic acid

(C) **STRANDEDNESS:** single

(D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:6:

GGTGATGTCC GGGTAGATCT CGGCGCAGCA CTCGTACTTC 40

CTGGTGTTGT AGGA 54

(2) **INFORMATION FOR SEQ ID NO:7:**

(i) **SEQUENCE CHARACTERISTICS:**

(A) **LENGTH:** 53 base pairs

(B) **TYPE:** nucleic acid

(C) **STRANDEDNESS:** single

(D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:7:

CCTACAACAC CAGGAAGTAC GAGTGCTGTG CCGAGATCTA 40

CCCGGACATC ACC 53

(2) **INFORMATION FOR SEQ ID NO:8:**

(i) **SEQUENCE CHARACTERISTICS:**

25

- (A) **LENGTH:** 53 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

5                   (xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:8:

CCTACAACAC CAGGAAGTAC GAGTGCTGAG CCGAGATCTA 40

CCCGGACATC ACC 53

(2) **INFORMATION FOR SEQ ID NO:9:**

(i) **SEQUENCE CHARACTERISTICS:**

- 10                   (A) **LENGTH:** 53 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:9:

15                   CCTACAACAC CAGGAAGTAC GAGTGCTGGG CCGAGATCTA 40

CCCGGACATC ACC 53

(2) **INFORMATION FOR SEQ ID NO:10:**

(i) **SEQUENCE CHARACTERISTICS:**

- 20                   (A) **LENGTH:** 53 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:10:

CCTACAACAC CAGGAAGTAC GAGTGCTGUG CCGAGATCTA 40

25                   CCCGGACATC ACC 53

(2) **INFORMATION FOR SEQ ID NO:11:**

(i) **SEQUENCE CHARACTERISTICS:**

- 30                   (A) **LENGTH:** 20 base pairs
- (B) **TYPE:** nucleic acid
- (C) **STRANDEDNESS:** single
- (D) **TOPOLOGY:** linear

26

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:11:

GGCGAGTGGG TCATCGTGGA 20

(2) **INFORMATION FOR SEQ ID NO:12:**

(i) **SEQUENCE CHARACTERISTICS:**

- 5 (A) **LENGTH:** 20 base pairs  
(B) **TYPE:** nucleic acid  
(C) **STRANDEDNESS:** single  
(D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:12:

10 GTGTTGTAGG TGCCACGGC 20

(2) **INFORMATION FOR SEQ ID NO:13:**

(i) **SEQUENCE CHARACTERISTICS:**

- 15 (A) **LENGTH:** 32 base pairs  
(B) **TYPE:** nucleic acid  
(C) **STRANDEDNESS:** single  
(D) **TOPOLOGY:** linear

(xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:13:

ACGTACCTTT TTTTAAAAAA AAAAAATAGG TC 32

(2) **INFORMATION FOR SEQ ID NO:14:**

(i) **SEQUENCE CHARACTERISTICS:**

- 20 (A) **LENGTH:** 29 base pairs  
(B) **TYPE:** nucleic acid  
(C) **STRANDEDNESS:** single  
(D) **TOPOLOGY:** linear

25 (xi) **SEQUENCE DESCRIPTION:** SEQ ID NO:14:

GGCTTCTTTG TCAAGATCAG CAAGAAGCA 29

## CLAIMS

We claim:

1. A method of detecting inheritance by an individual of a selected genetic disease or disorder associated with a polymorphic site comprising a base substitution or a base deletion or insertion, comprising the steps of:

(a) collecting a genomic DNA sample from said individual;

(b) amplifying a selected portion of said genomic DNA sample by polymerase chain reaction to produce an amplified DNA segment containing said polymorphic site and having a size such that a difference in mass between said amplified DNA segment and a corresponding portion of a reference DNA sample can be resolved by electrospray ionization mass spectrometry;

(c) desalting said amplified DNA segment to produce a desalted amplified DNA segment;

(d) determining the mass of said desalted amplified DNA segment by electrospray ionization mass spectrometry; and

(e) comparing said mass to a reference mass determined for said corresponding portion of said reference DNA sample, wherein a difference between said mass and said reference mass indicates inheritance of said genetic disease or disorder.

2. The method of claim 1 wherein said electrospray ionization mass spectrometry comprises ion detection with a quadrupole detector.

3. The method of claim 1 wherein said electrospray ionization mass spectrometry comprises ion detection with a Fourier transform ion cyclotron resonance mass detector.

4. The method of claim 1 wherein the size of the amplified DNA segment is no greater than about 55 base pairs.

5 5. The method of claim 4 wherein the size of the amplified DNA segment is no greater than about 45 base pairs.

6. The method of claim 1 wherein the desalting comprises high pressure liquid chromatography.

10 7. The method of claim 1 wherein the desalting comprises molecular weight cut-off spin filtration.

8. The method of claim 1 wherein the desalting comprises alcohol precipitation.

15 9. The method of claim 1 wherein said selected genetic disease or disorder is benign neonatal familial convulsions.

10. The method of claim 9 wherein said amplifying comprises polymerase-catalyzed extension of primers having sequences represented as SEQ ID NO:11 and SEQ ID NO:12.

20 11. The method of claim 1 wherein said selected genetic disease or disorder is attenuated polyposis coli.

25 12. The method of claim 11 wherein said amplifying comprises polymerase-catalyzed extension of primers having sequences represented as SEQ ID NO:13 and SEQ ID NO:14.

13. A method of detecting heterozygosity at a polymorphic site in the genome of an individual,

wherein said polymorphic site comprises a site at which a base substitution or base deletion or insertion occurs in a mutant allele as compared to a wild type allele, comprising the steps of:

5 (a) collecting a genomic DNA sample from said individual;

(b) amplifying a selected portion of said genomic DNA sample by polymerase chain reaction to produce first and second amplified DNA segments representing corresponding portions of both homologous chromosomes containing said polymorphic site, wherein said first and second amplified DNA segments each have a size such that a difference in mass between said first and second amplified DNA segments can be resolved by electrospray ionization mass spectrometry;

10 (c) desalting said first and second amplified DNA segments to produce respective first and second desalted amplified DNA segments;

(d) determining the masses of said first and second desalted amplified DNA segments by electrospray ionization mass spectrometry; and

15 (e) comparing said masses of said first and second desalted amplified DNA segments, wherein a difference in said masses indicates heterozygosity at said polymorphic site.

20 14. The method of claim 13 wherein said electrospray ionization mass spectrometry comprises ion detection with a quadrupole detector.

30 15. The method of claim 13 wherein said electrospray ionization mass spectrometry comprises ion detection with a Fourier transform ion cyclotron resonance mass detector.

16. The method of claim 13 wherein the size of the first and second amplified DNA segments is no greater than about 55 base pairs.

17. The method of claim 16 wherein the size of the first and second amplified DNA segments is no greater than about 45 base pairs.

18. The method of claim 13 wherein the desalting comprises high pressure liquid chromatography.

19. The method of claim 13 wherein the desalting comprises molecular weight cut-off spin filtration.

20. The method of claim 13 wherein the desalting comprises alcohol precipitation.

21. A method of detecting a polymorphism at a polymorphic site comprising a base substitution or a base deletion or insertion, comprising the steps of:

- (a) collecting a DNA sample to be tested;
- (b) amplifying a selected portion of said DNA sample by polymerase chain reaction to produce an amplified DNA segment containing said polymorphic site and having a size such that a difference in mass between said amplified DNA segment and a corresponding portion of a reference DNA sample can be resolved by electrospray ionization mass spectrometry;
- (c) desalting said amplified DNA segment to produce a desalted amplified DNA segment;
- (d) determining the mass of said desalted amplified DNA segment by electrospray ionization mass spectrometry; and
- (e) comparing said mass to a reference mass determined for said corresponding portion of said reference DNA sample, wherein a difference between



said mass and said reference mass indicates the presence of the polymorphism.

22. The method of claim 21 wherein said electrospray ionization mass spectrometry comprises ion detection with a quadrupole detector.

23. The method of claim 21 wherein said electrospray ionization mass spectrometry comprises ion detection with a Fourier transform ion cyclotron resonance mass detector.

24. The method of claim 21 wherein the size of the amplified DNA segment is no greater than about 55 base pairs.

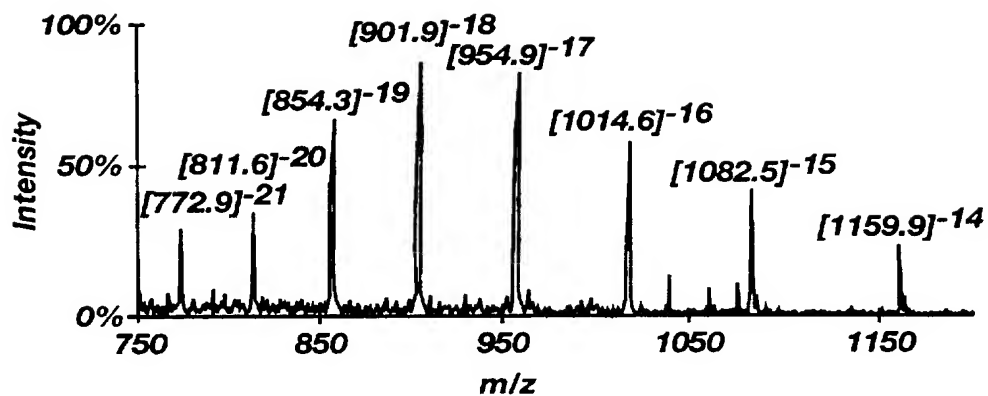
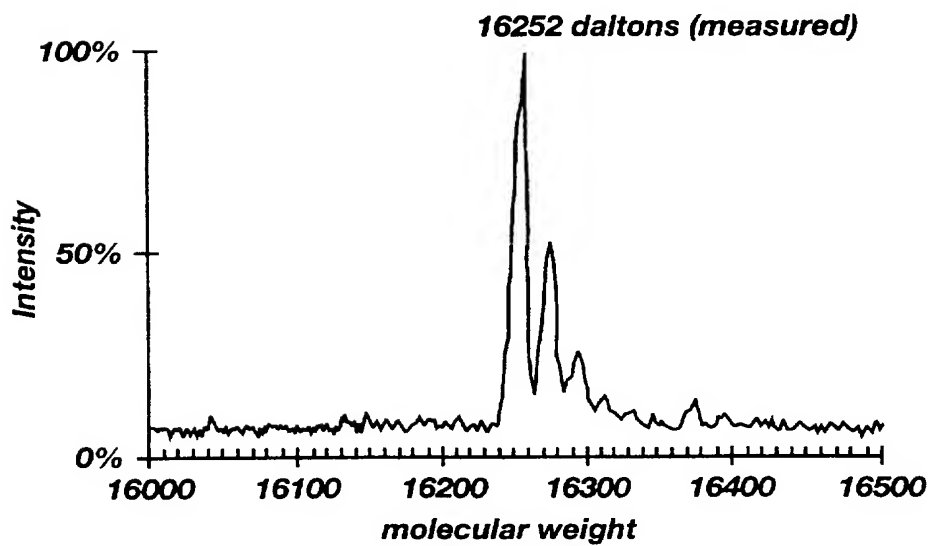
25. The method of claim 24 wherein the size of the amplified DNA segment is no greater than about 45 base pairs.

26. The method of claim 21 wherein the desalting comprises high pressure liquid chromatography.

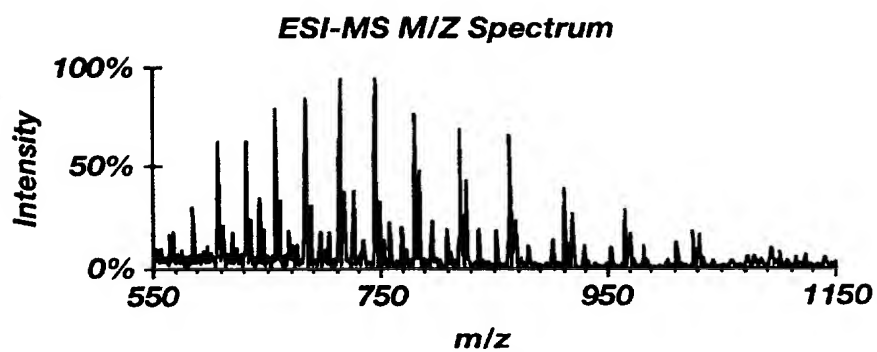
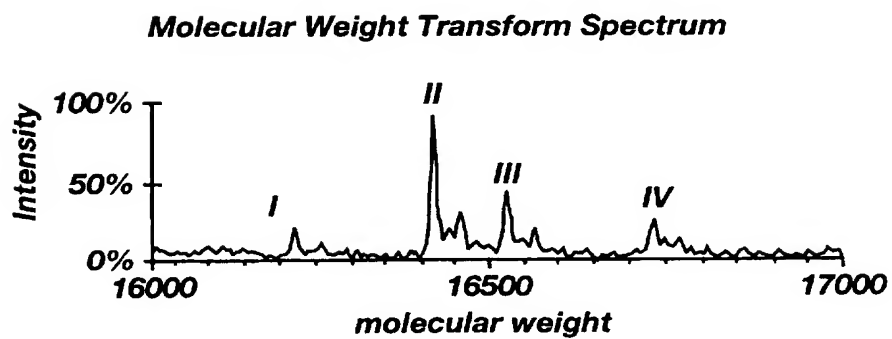
27. The method of claim 21 wherein the desalting comprises molecular weight cut-off spin filtration.

28. The method of claim 21 wherein the desalting comprises alcohol precipitation.

1/6

**Fig. 1****Fig. 2**

2/6

**Fig. 3****Fig. 4**

SUBSTITUTE SHEET (RULE 26)

3/6

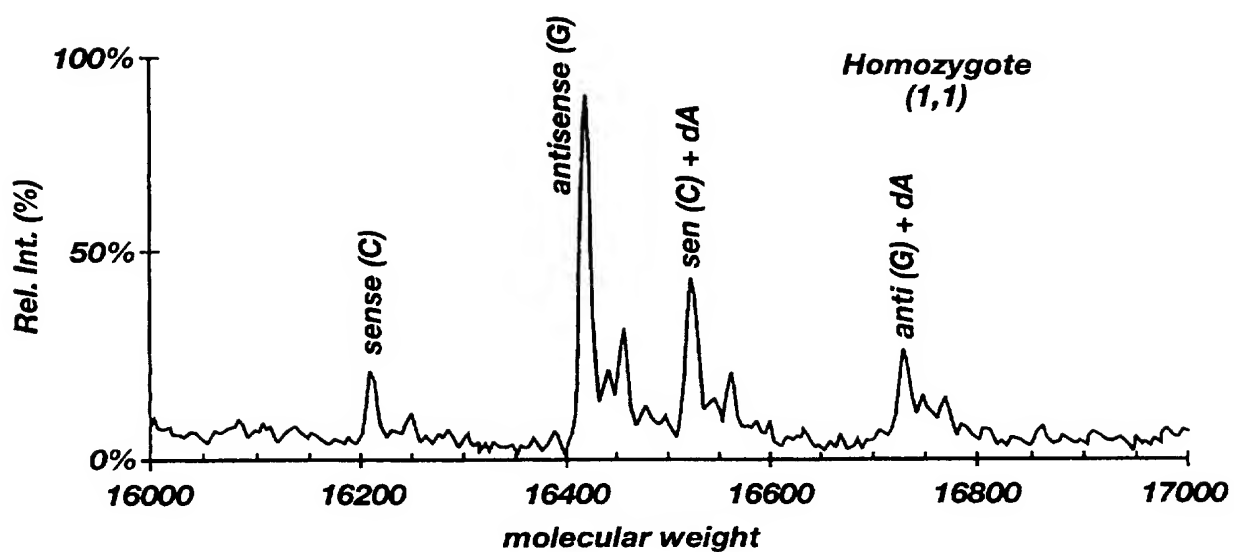


Fig. 5A

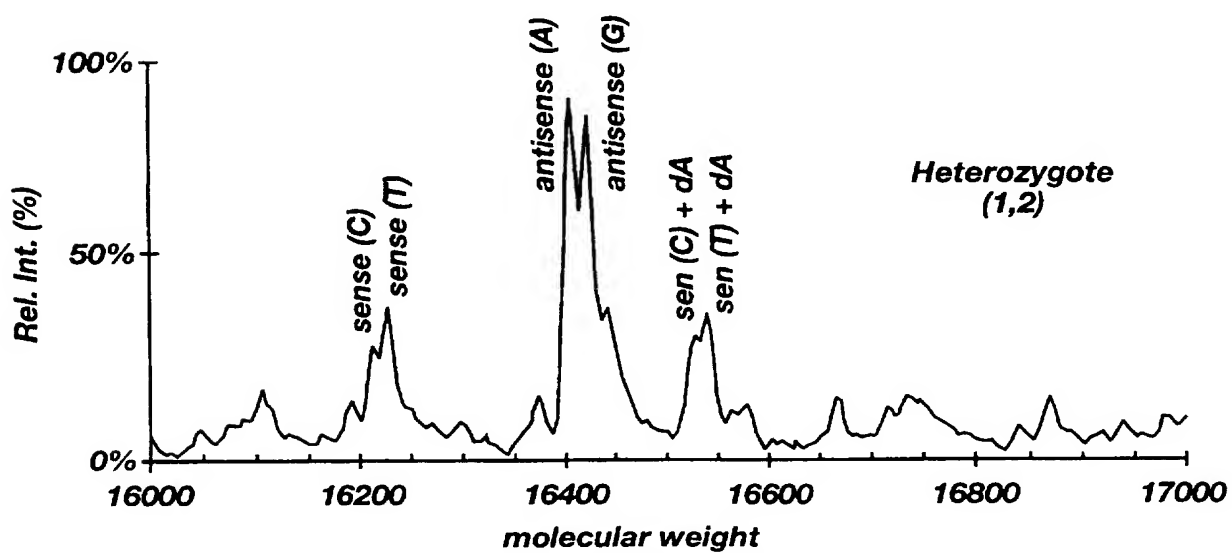
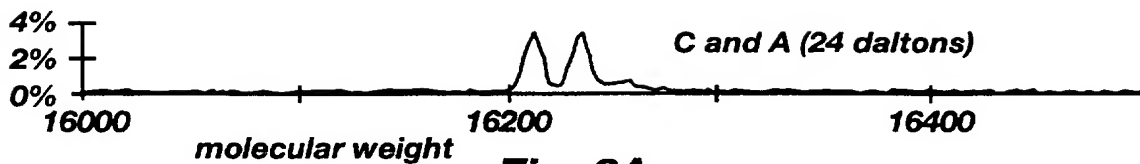
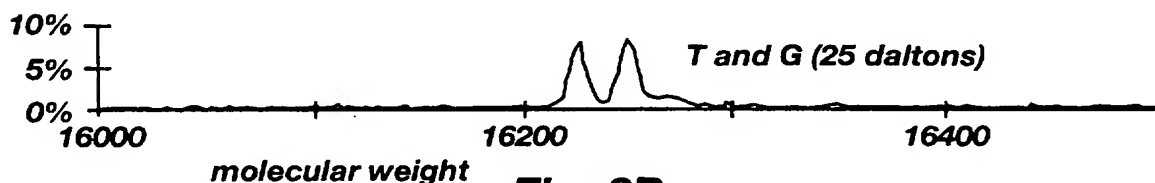


Fig. 5B

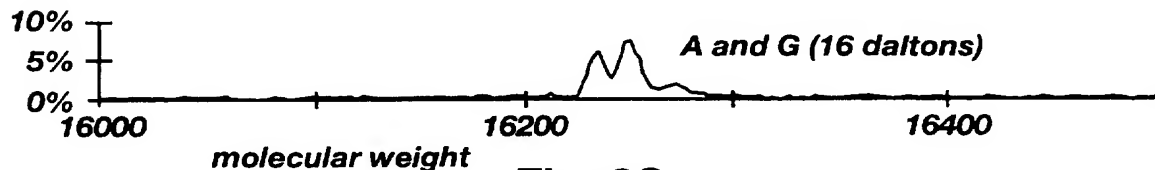
4/6



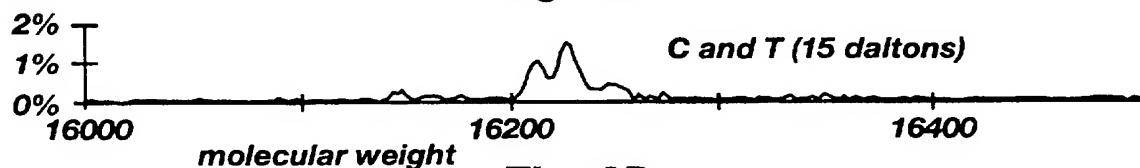
**Fig. 6A**



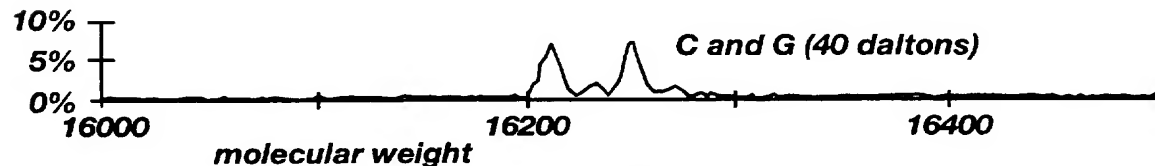
**Fig. 6B**



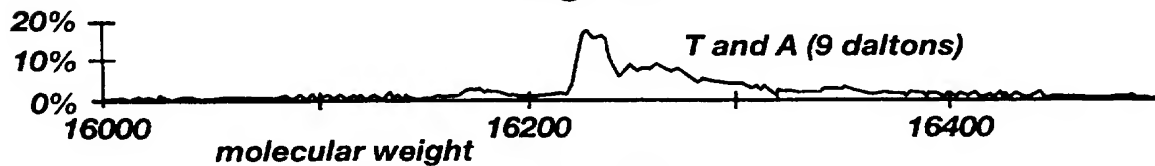
**Fig. 6C**



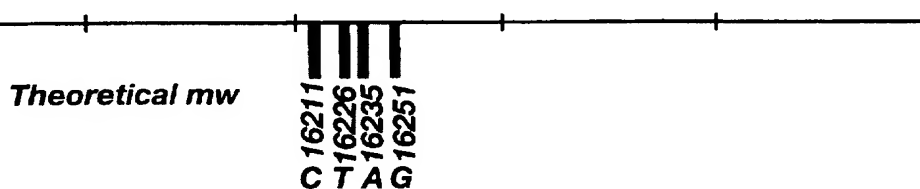
**Fig. 6D**



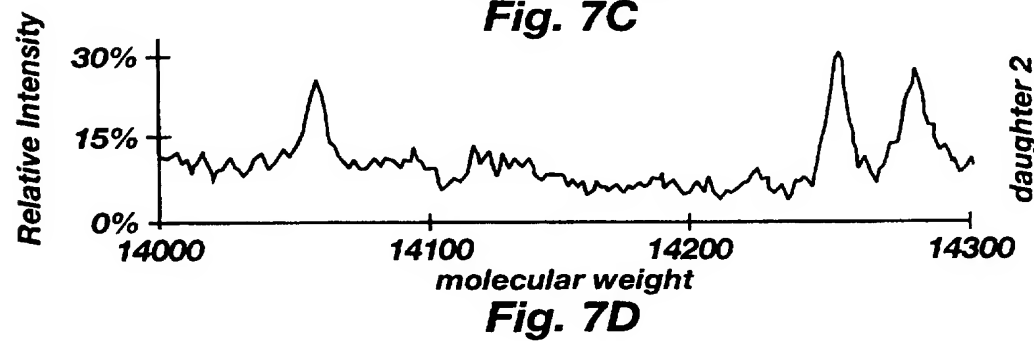
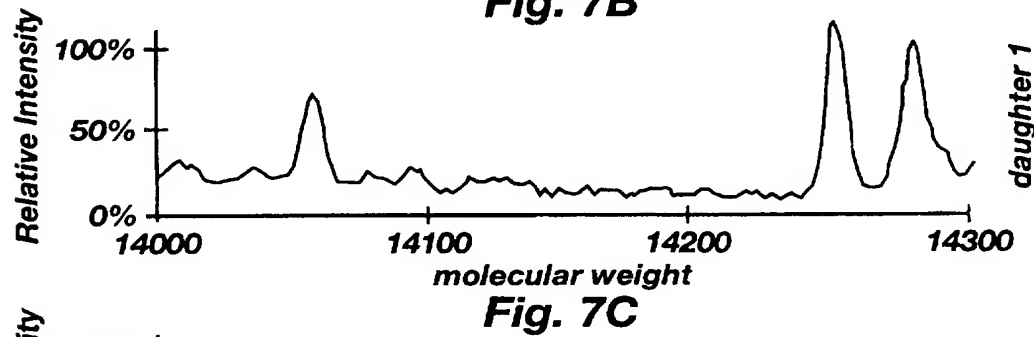
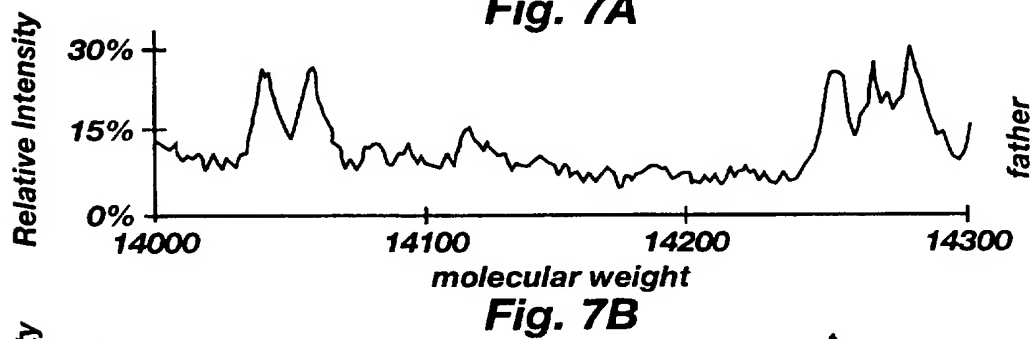
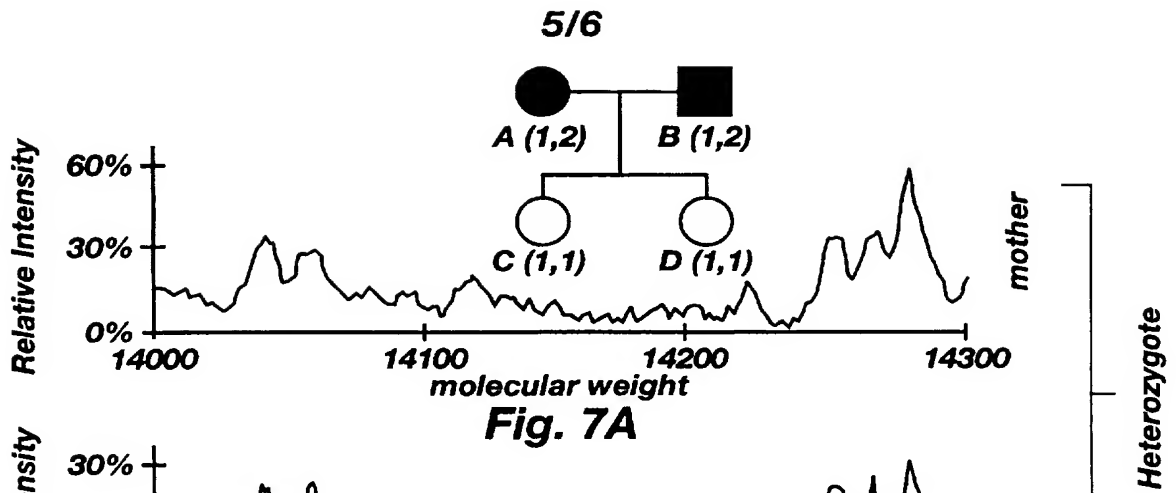
**Fig. 6E**



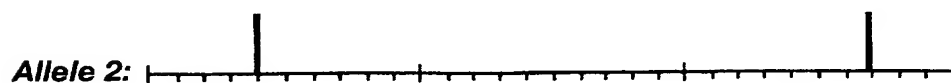
**Fig. 6F**



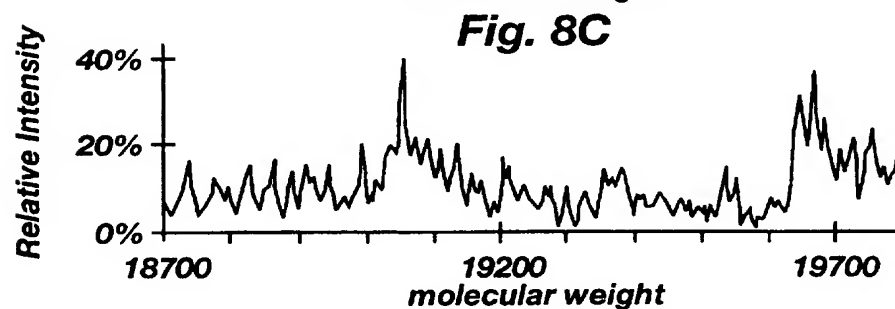
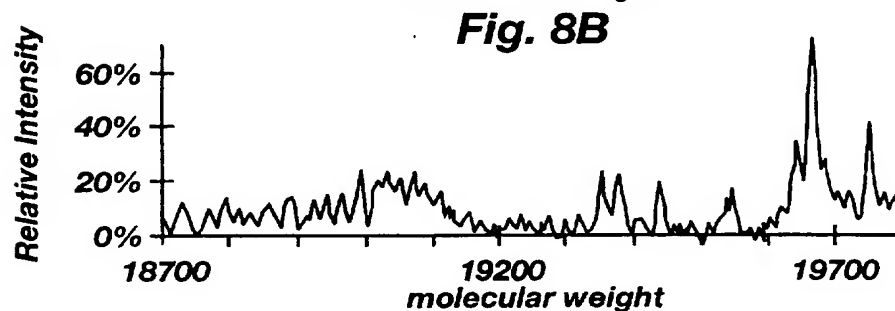
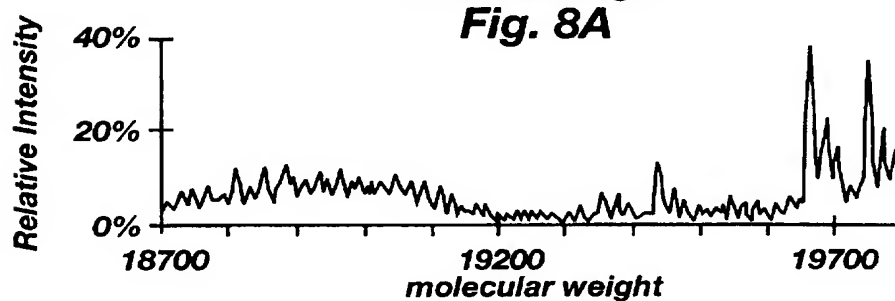
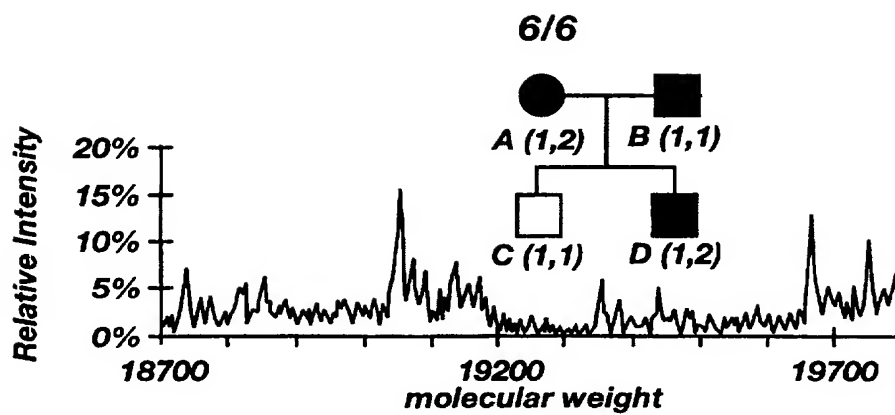
**Fig. 6G**



**Fig. 7E**



**Fig. 7F**

**Fig. 8E****Fig. 8F**

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/08518

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :C12Q 1/68; C12P 19/34

US CL :435/6, 91.2

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
NONEElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Extra Sheet.**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHAUDHARY et al. Characterization of an $N^6$ -Oxopropenyl-2'-Deoxyadenosine Adduct in Malondialdehyde-Modified DNA Using Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry. Carcinogenesis. 1996, Vol. 17, No. 5, pages 1167-1170, especially pages 1168-1169.	1-28
Y	HETTICH et al. MALDI-FTMS for the Characterization of Ultraviolet and X-Ray Damage to Nucleic Acid Constituents. Proceedings of the 41st ASMA Conference on Mass Spectrometry and Allied Topics. June 1993, pages 250A and 250B, especially page 250A.	1-28

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

08 SEPTEMBER 1997

Date of mailing of the international search report

07 OCT 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

Carla Myers

Telephone No. (703) 308-0196



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/08518

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

<b>Category*</b>	<b>Citation of document, with indication, where appropriate, of the relevant passages</b>	<b>Relevant to claim No.</b>
Y	ZHAO et al. Capillary Isotachopheresis-Mass Spectrometric Determination of DNA Damage. Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics. 1995, page 592, see entire document.	1-28
Y	US 5,221,518 A (R. L. MILLS) 22 June 1993, col. 44, 54-56, and 64.	1-28
Y	DELGADO-ESCUETA et al. Progress in Mapping Epilepsy Genes. Epilepsia. 1994, Vol. 35, Supplement 1, pages S29-S40, especially page S35.	9, 10
Y	DOBBIE et al. Mutational Analysis of the First 14 Exons of the Adenomatous Polyposis Coli (APC) Gene. European Journal of Cancer. 1994, Vol. 30A, No. 11, pages 1709-1713, especially page 1710.	11, 12

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/08518

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS; DIALOG: MEDLINE, CA, DERWENT PATENTS, EMBASE, BIOSIS, SCISEARCH

search terms: ESI-MS, electrospray ionization mass spectrometry, fourier transform ion cyclotron resonance, DNA, mutation, polymorphism, polynucleotide, nucleic acid, benign neonatal familial convulsions, attenuated polyposis coli.